MIT
International Center for
Air Transportation

Trajectory Clustering and Classification for Characterization of Air Traffic Flows

Mayara Condé Rocha Murça
Prof. John Hansman

JUP Meeting
April 21, 2016

Motivation

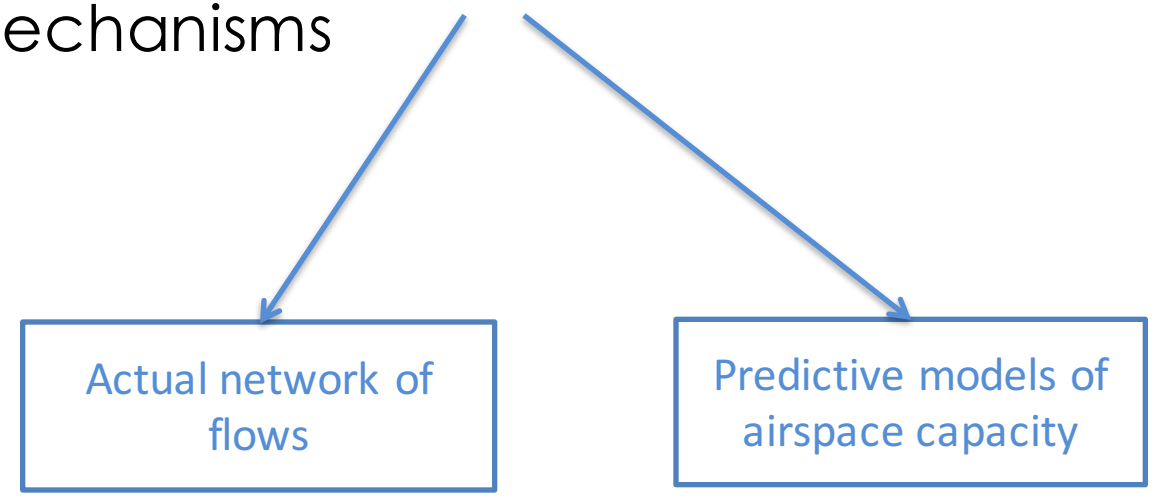
- The Air Traffic Management System is characterized by a highly computerized environment in which massive amounts of data are generated daily as planning/operations occur
- Current system capabilities do not take full advantage of this big data potential
  - Example: Current traffic flow management initiatives rely on limited empirical estimates of airport capacity (eg.: that do not account for weather impacts in the terminal/transition airspace)
- Future operations can benefit from big data analytics tools that provide/enhance the following capabilities:
  - Post-event efficiency assessment
  - Monitoring and alerting
  - Real time decision support

Research Goal

- To develop an analytics framework for characterizing historical trajectory patterns in the airspace that can be used to evaluate the performance of past operations and generate inputs for air traffic flow management mechanisms

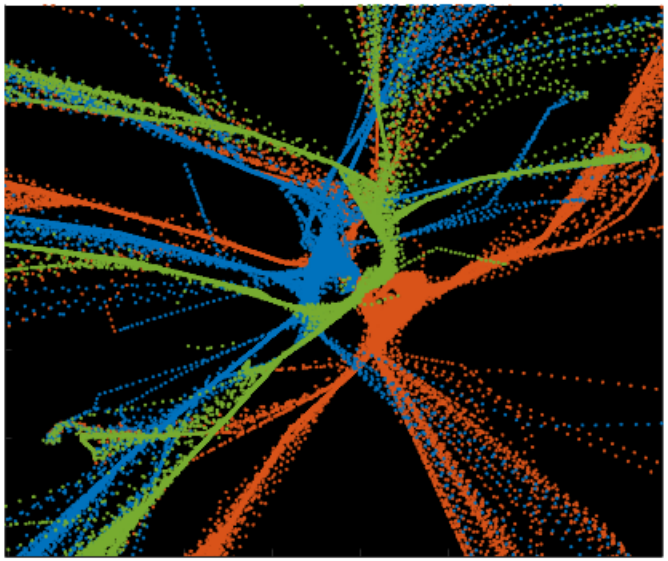
Research Goal

- To develop an analytics framework for characterizing historical trajectory patterns in the airspace that can be used to evaluate the performance of past operations and generate **inputs** for air traffic flow management mechanisms



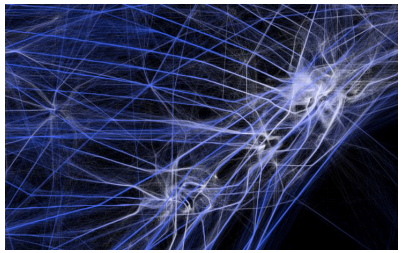
Approach

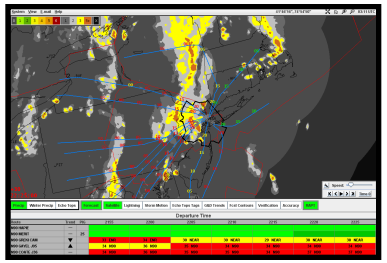
- Development of a comprehensive data mining framework based on recorded radar tracks and weather impact measures in order to:
  - Identify major flight trajectory patterns
  - Assess the conformance of flight trajectories with respect to identified patterns
  - Identify and characterize patterns of airspace use and associated causes
- Application to an initial case study
  - Tactical ATC operations in the transition/terminal airspace for the New York Metro (JFK, EWR and LGA)



Dataset

- Sources and scope
  - Flight trajectory data
    - ETMS radar tracks: one minute updates of aircraft state in the domestic airspace
  - Weather impact measures
    - En-route convective impacts: hourly blockage status for NY departure routes from the Route Availability Planning Tool (RAPT)
    - Winds, ceiling and visibility: hourly airport weather report from the Aviation System Performance Metrics (ASPM) database





- Time period
  - 1st phase: 70 days 2013-2015
  - 2nd phase: 1000 days 2013-2015

Methodology

RAW DATA:
OBSERVED FLIGHT TRACKS

FILTERED AND RESAMPLED
FLIGHT TRACKS

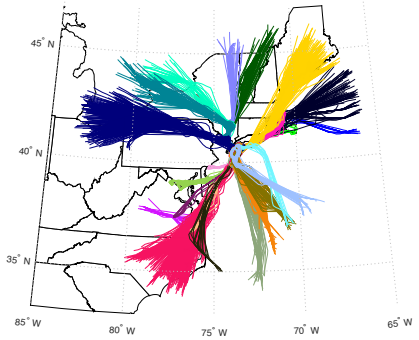
RAW DATA:
WEATHER IMPACT MEASURES

CORRELATION
ANALYSIS

RESOURCE IDENTIFICATION

- **CLUSTERING** of trajectories using **DBSCAN** (Density-based spatial clustering of applications with noise) algorithm

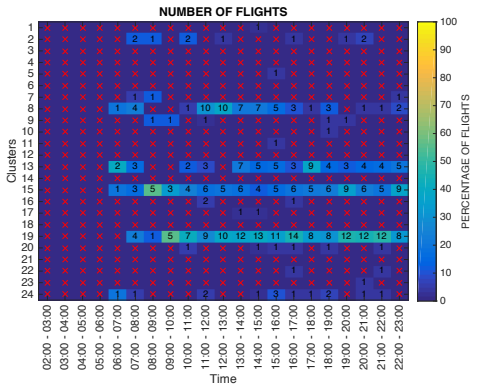
Airspace Structure



RESOURCE USE AND NON-CONFORMITY DETECTION

- **CLASSIFICATION** of trajectories using **RANDOM FORESTS** and **CONFORMAL PREDICTION**

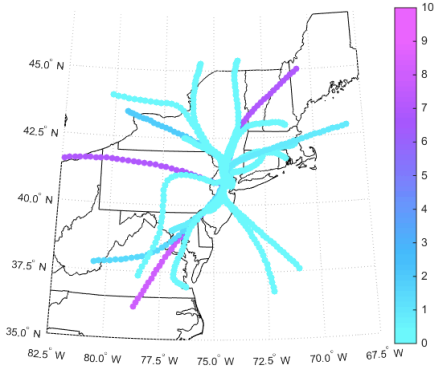
Hourly Resource Use Vectors
Daily Resource Use Matrices



OPERATIONAL PATTERNS

- **CLUSTERING** of Resource Use Vectors using **K-MEANS** algorithm

Hourly Resource Use Patterns



Methodology

RAW DATA:
OBSERVED FLIGHT TRACKS

FILTERED AND RESAMPLED
FLIGHT TRACKS

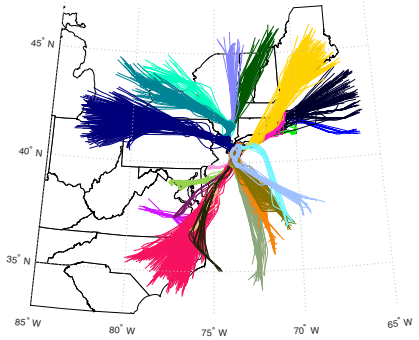
RAW DATA:
WEATHER IMPACT MEASURES

CORRELATION
ANALYSIS

RESOURCE IDENTIFICATION

- **CLUSTERING** of trajectories using **DBSCAN** (Density-based spatial clustering of applications with noise) algorithm

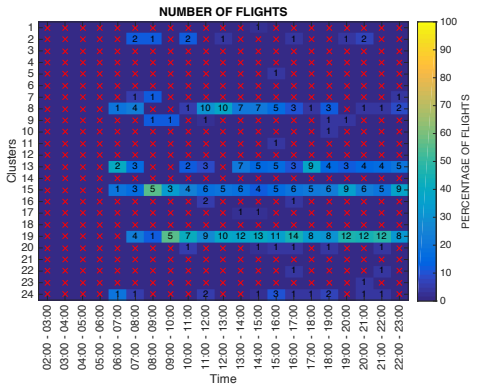
Airspace Structure



RESOURCE USE AND NON-CONFORMITY DETECTION

- **CLASSIFICATION** of trajectories using **RANDOM FORESTS** and **CONFORMAL PREDICTION**

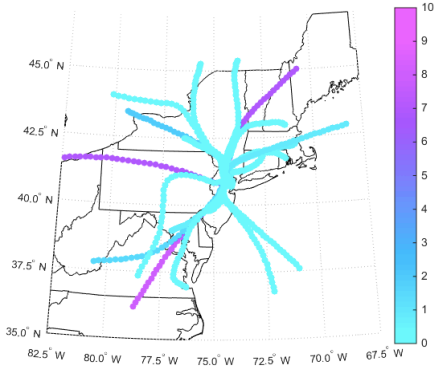
Hourly Resource Use Vectors
Daily Resource Use Matrices



OPERATIONAL PATTERNS

- **CLUSTERING** of Resource Use Vectors using **K-MEANS** algorithm

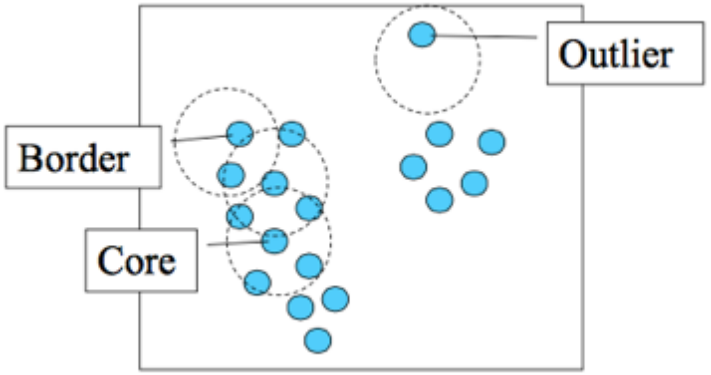
Hourly Resource Use Patterns



Clustering Analysis: DBSCAN

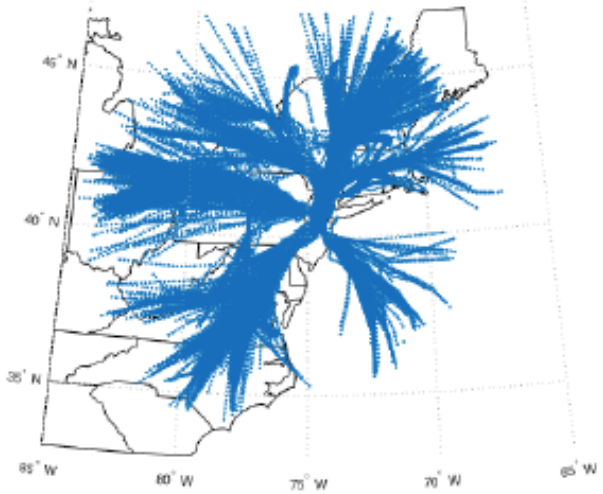
- Concept: a cluster is determined by a set of density-connected points in the data space (Ester et al., 1996)
- Two input parameters:
  - *MinPts*: Minimum number of points
  - *Epsilon*: Distance threshold
    - *Epsilon*–neighborhood: $N_{Eps}(p) = \{q \in D / dist(p,q) \leq \varepsilon\}$

Core point: It contains more than *MinPts* in its *Epsilon*-neighborhood
Border point: It is density-reachable from a core point
Noise point (outlier): It is not density-reachable from any other point in the database

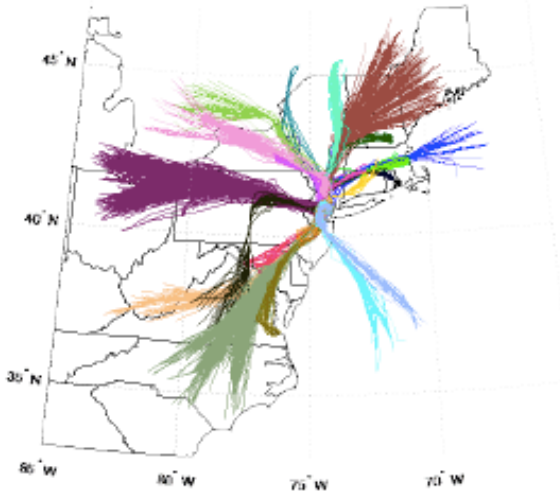


Resource Identification

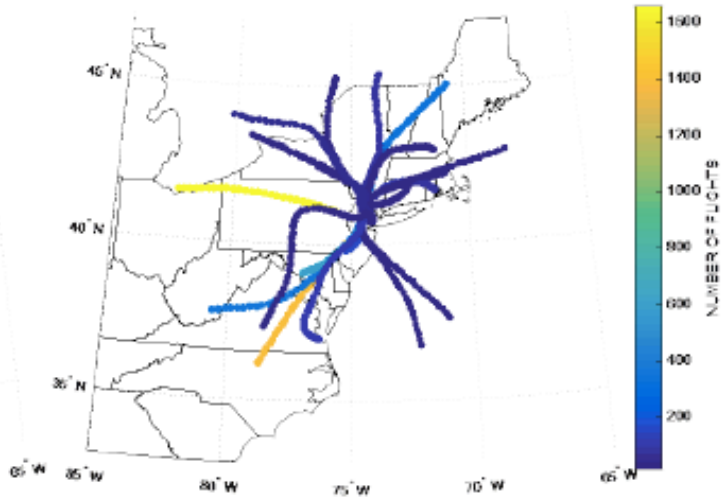
Identification of EWR Arrival Flows



13 days of arrivals…



…23 clusters…



…23 Arrival Resources…

- Clustering algorithm parameters determined with sensitivity analysis and cluster validity indices evaluation
- Resulting clusters captured ~92% of all trajectories (in other words, 8% are non-conforming trajectories that do not fit to any of the identified clusters)

Methodology

RAW DATA:
OBSERVED FLIGHT TRACKS

FILTERED AND RESAMPLED
FLIGHT TRACKS

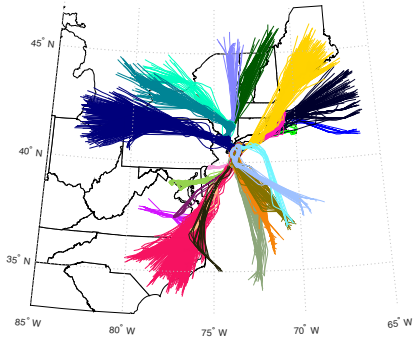
RAW DATA:
WEATHER IMPACT MEASURES

CORRELATION
ANALYSIS

RESOURCE IDENTIFICATION

- **CLUSTERING** of trajectories using **DBSCAN** (Density-based spatial clustering of applications with noise) algorithm

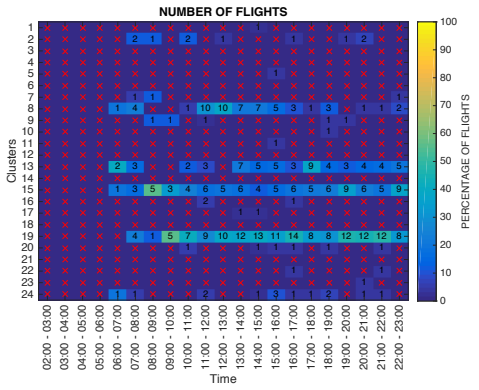
Airspace Structure



RESOURCE USE AND NON-CONFORMITY DETECTION

- **CLASSIFICATION** of trajectories using **RANDOM FORESTS** and **CONFORMAL PREDICTION**

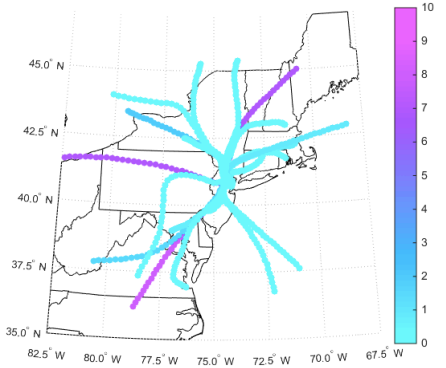
Hourly Resource Use Vectors
Daily Resource Use Matrices



OPERATIONAL PATTERNS

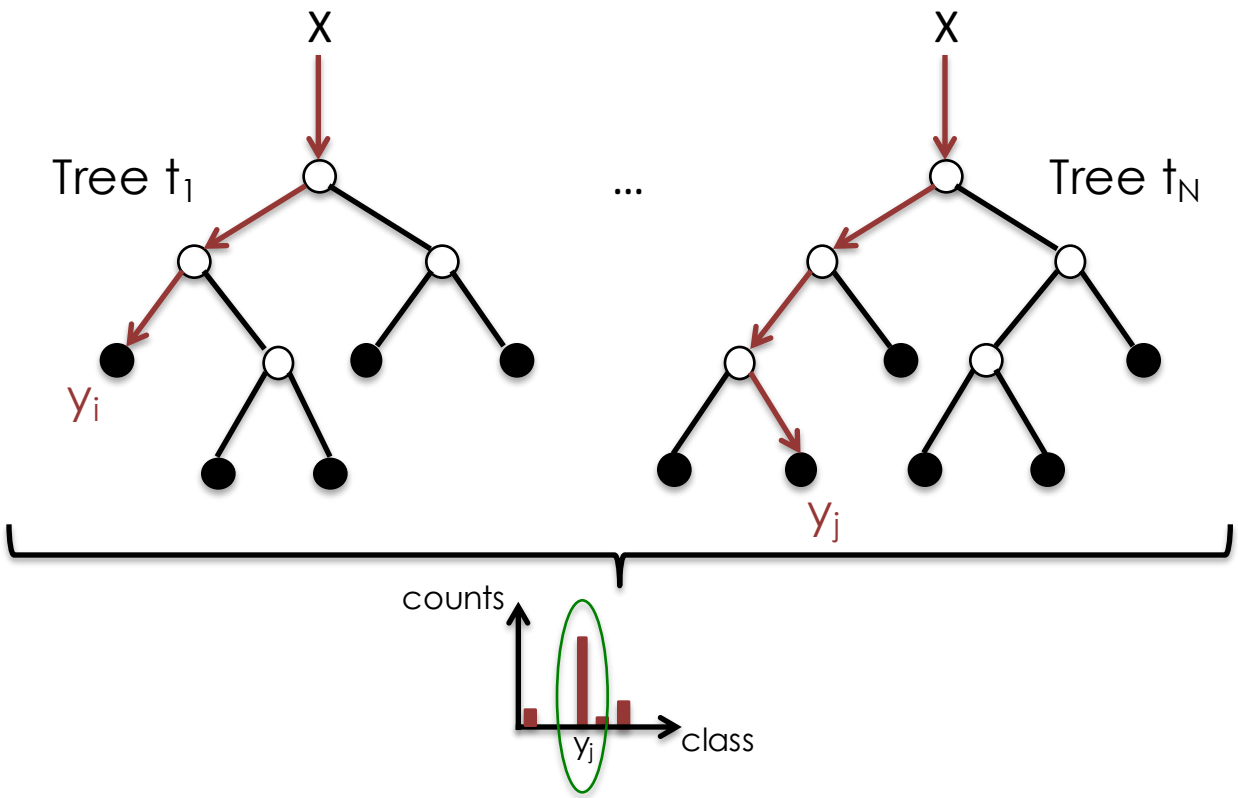
- **CLUSTERING** of Resource Use Vectors using **K-MEANS** algorithm

Hourly Resource Use Patterns



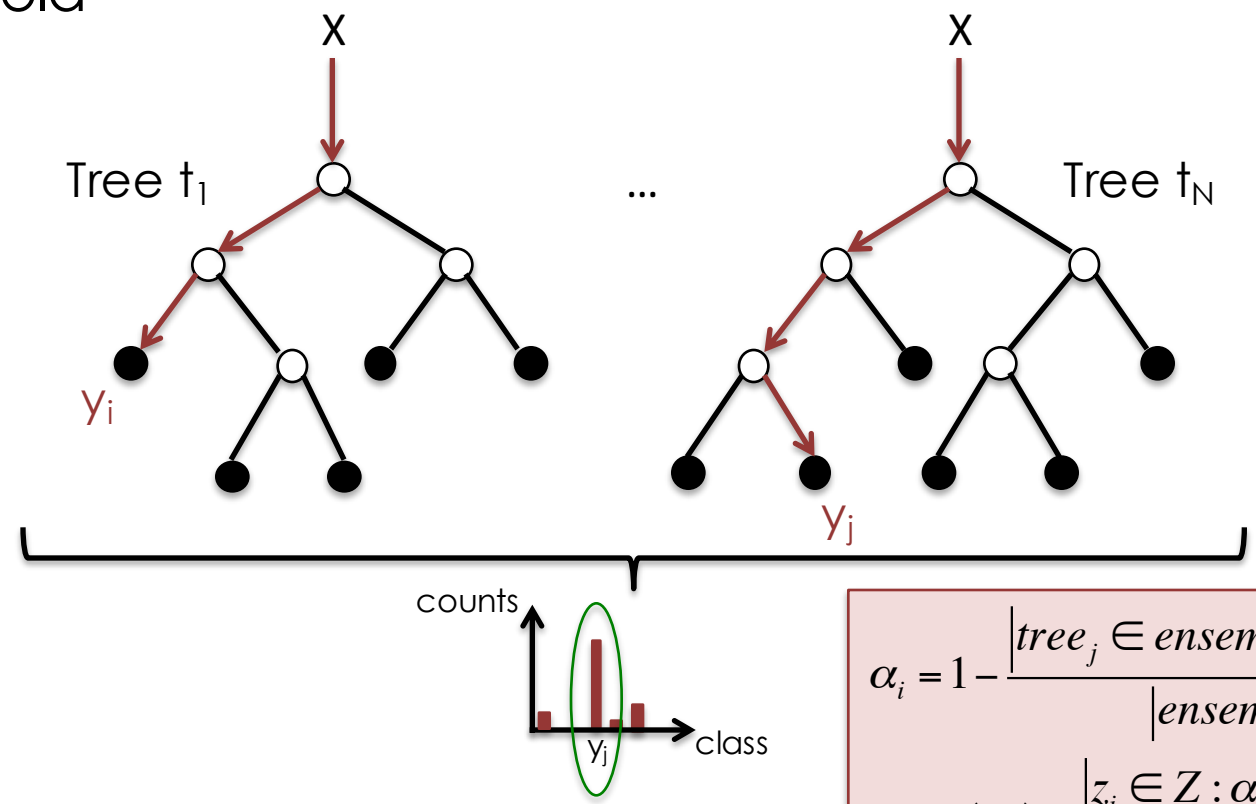
Classification Scheme
Random Forests and Conformal Prediction

- Random Forests determines the class of a new observation by the majority of votes from an ensemble of decision trees created from bootstrap samples of the data (Breiman, 2001)



Classification Scheme
Random Forests and Conformal Prediction

- Conformal Prediction generates a confidence measure *p* for the prediction (Shaffer and Vovk, 2008; Bhattacharyya, 2013)
- Non-conforming behaviors identified when *p* is less than a threshold



$$\alpha_i = 1 - \frac{\left|tree_j \in ensemble : y_i = \hat{y}_i^j\right|}{\left|ensemble\right|}$$

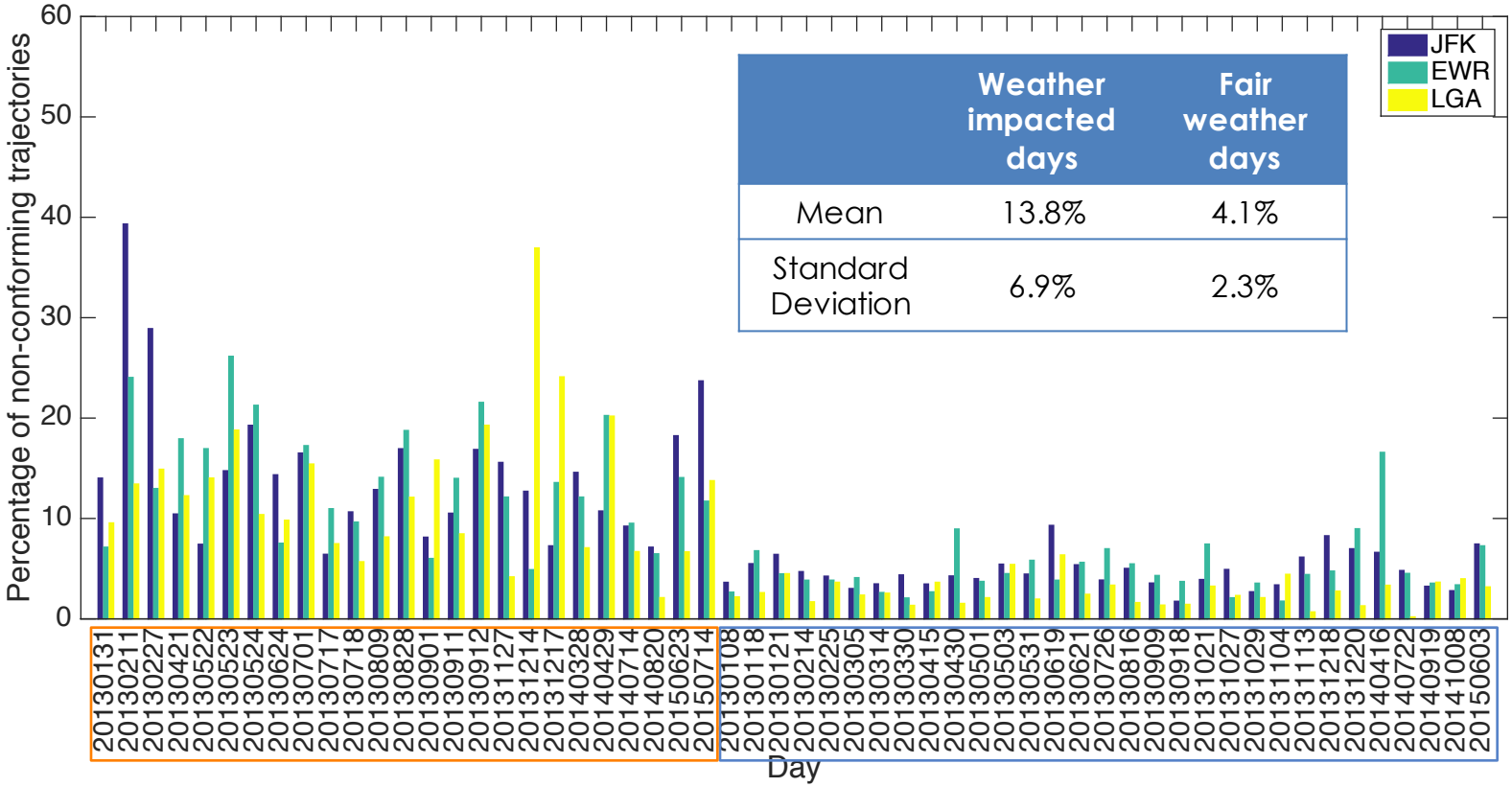
$$p\left(\hat{y}_n\right) = \frac{\left|z_i \in Z : \alpha_i \geq \alpha_n^{\hat{y}_n}\right| + 1}{\left|Z\right| + 1}$$

Classification Scheme
Random Forests and Conformal Prediction

- Classification performance assessed with 5-fold cross validation
  - Resource assignment accuracy: > 98%
  - Non-conformity detection accuracy: > 93%

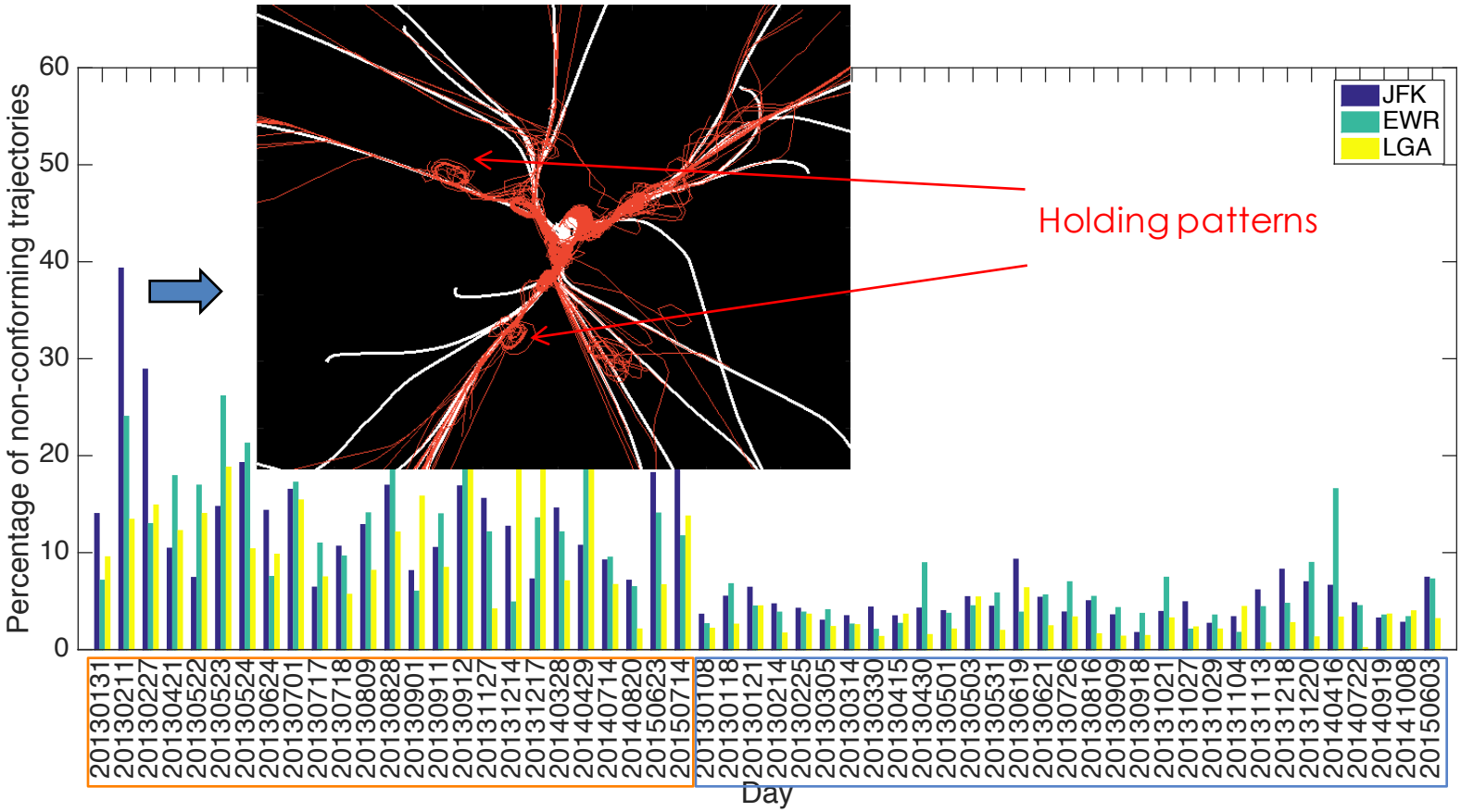
Flight Trajectory Conformance

- Identification of non-conforming behaviors for arrival trajectories



Flight Trajectory Conformance

- Identification of non-conforming behaviors for arrival trajectories



Days with fair weather conditions

Days with weather impacts (convection or adverse winds / ceiling / visibility)

Resource Use Matrix

- Aggregation of classification results to generate a Resource Use Matrix (RUM) for each day of operations
  - A matrix is defined as a RUM if each element r_{ij} contains the number of trajectories that arrived/departed using pattern *i* during hour *j*
- The RUM replaces individual trajectory records and generates a compact representation of flows during the day that can be used for:
  - Pointwise comparisons of tactical operations
  - Large statistical analysis of airspace use

Methodology

RAW DATA:
OBSERVED FLIGHT TRACKS

FILTERED AND RESAMPLED
FLIGHT TRACKS

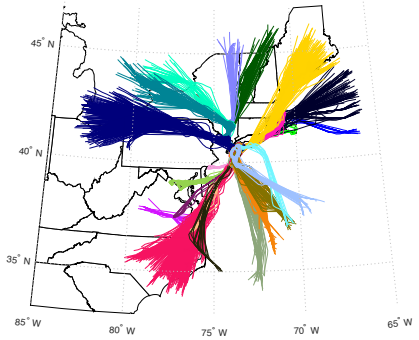
RAW DATA:
WEATHER IMPACT MEASURES

CORRELATION
ANALYSIS

RESOURCE IDENTIFICATION

- **CLUSTERING** of trajectories using **DBSCAN** (Density-based spatial clustering of applications with noise) algorithm

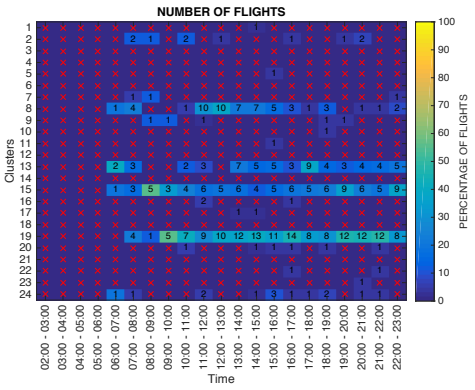
Airspace Structure



RESOURCE USE AND NON-CONFORMITY DETECTION

- **CLASSIFICATION** of trajectories using **RANDOM FORESTS** and **CONFORMAL PREDICTION**

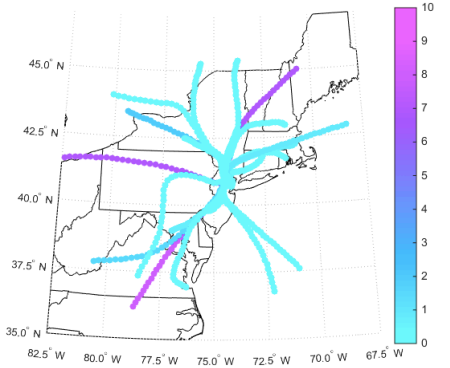
Hourly Resource Use Vectors
Daily Resource Use Matrices



OPERATIONAL PATTERNS

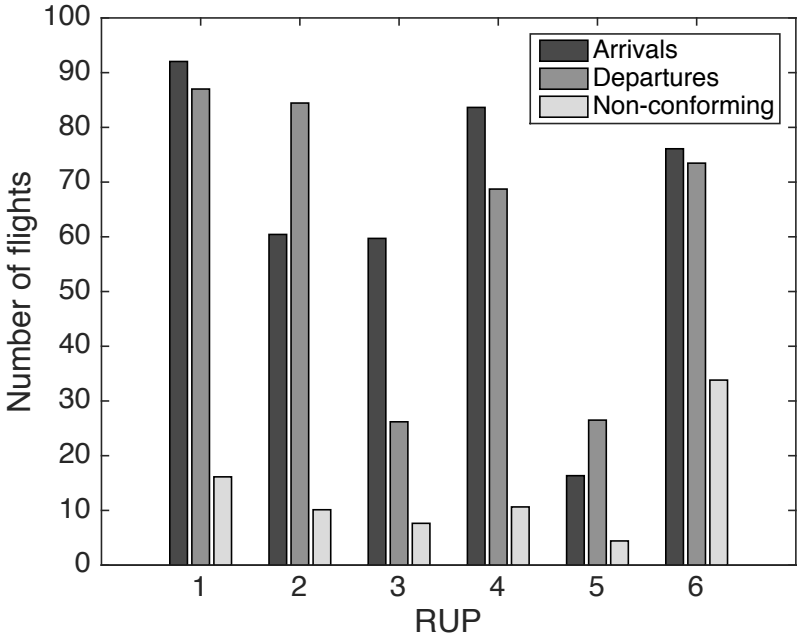
- **CLUSTERING** of Resource Use Vectors using **K-MEANS** algorithm

Hourly Resource Use Patterns



Resource Use Patterns

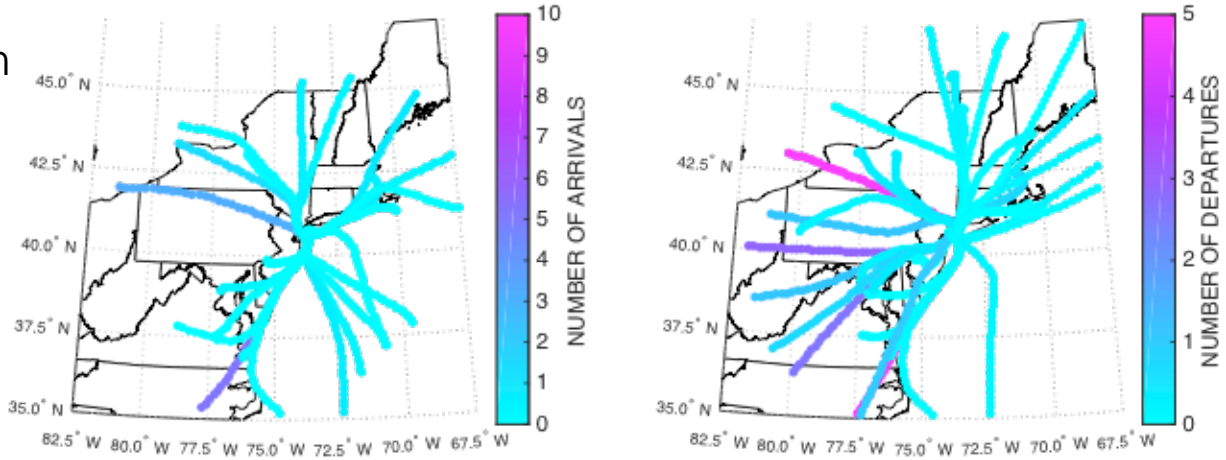
- Six aggregate modes of operation (Resource Use Patterns – RUP) were identified for the NY Metro
  - RUP 1: High arr/dep throughput
  - RUP 2: High dep, medium arr throughput
  - RUP 3: Medium arr, low dep throughput
  - RUP 4: High arr, medium dep throughput
  - RUP 5: Low arr/dep throughput
  - RUP 6: Medium arr/dep throughput, high non-conformance



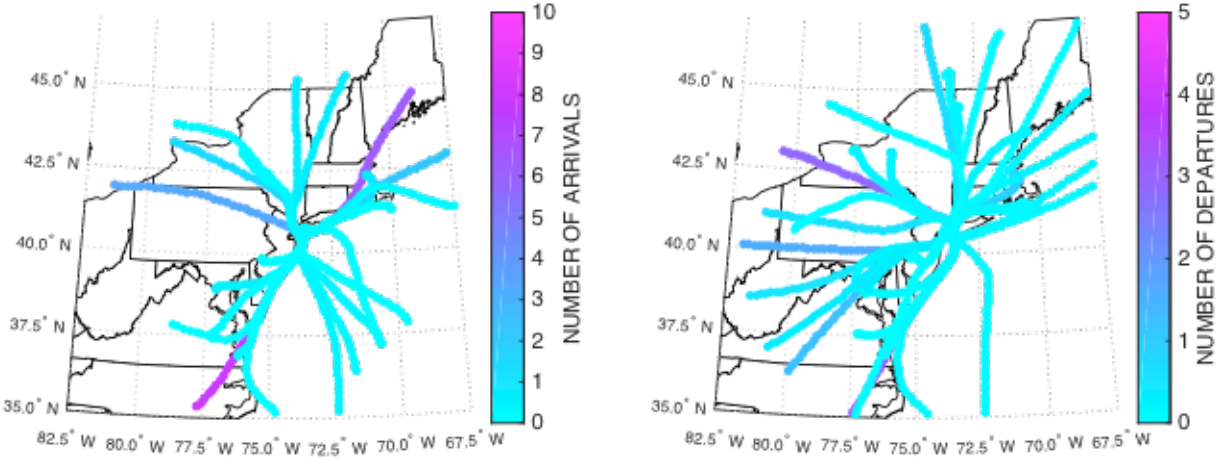
Resource Use Patterns

- Six aggregate modes of operation (Resource Use Patterns – RUP) were identified for the NY Metro

RUP 2: High dep, medium arr throughput

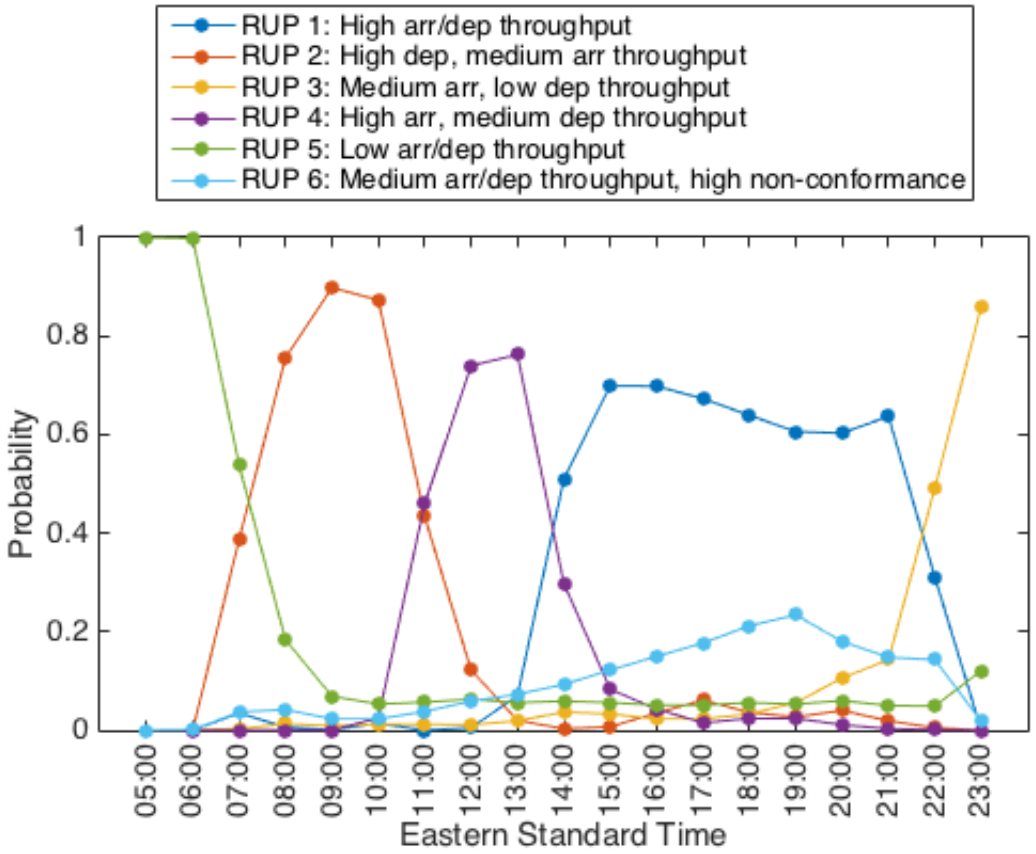


RUP 4: High arr, medium dep throughput



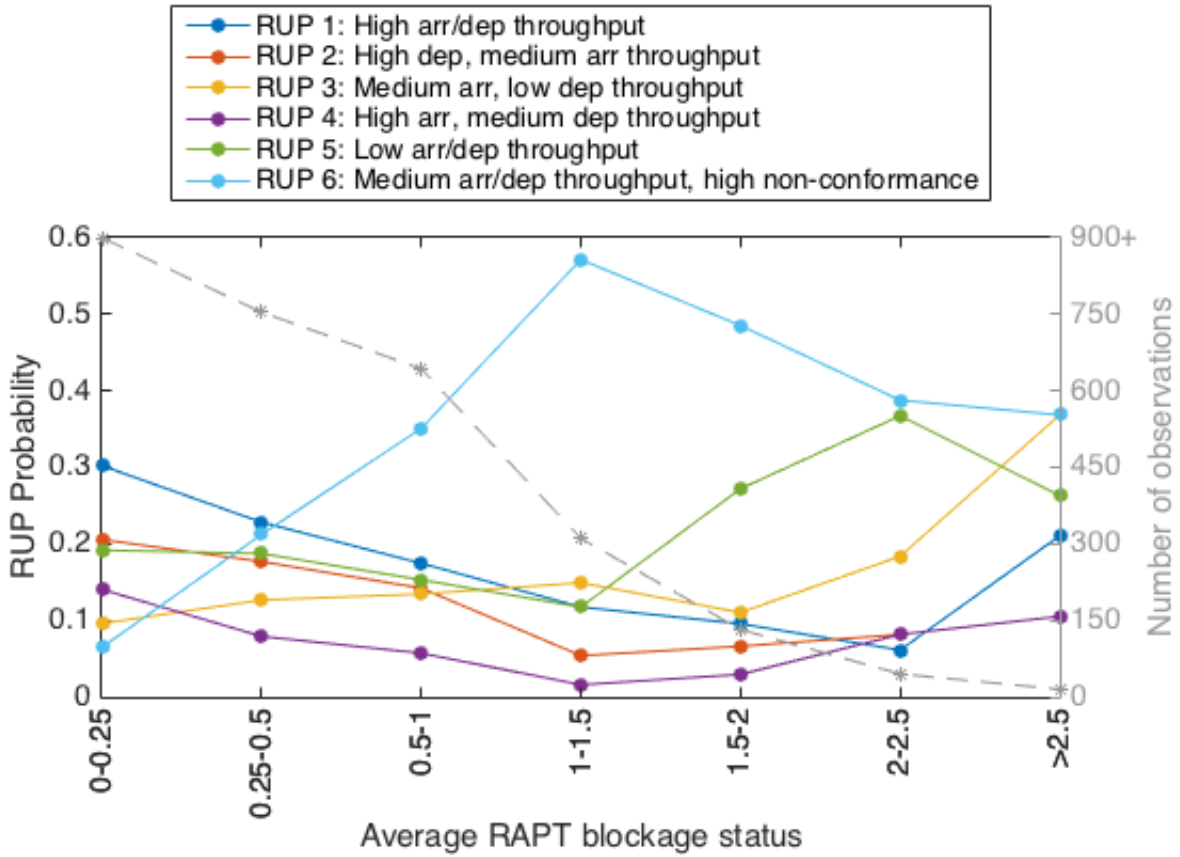
Resource Use Patterns

- Patterns are correlated with time of day and constraints in the system
  - Probability of RUP occurrence by hour reveals major demand patterns during the day



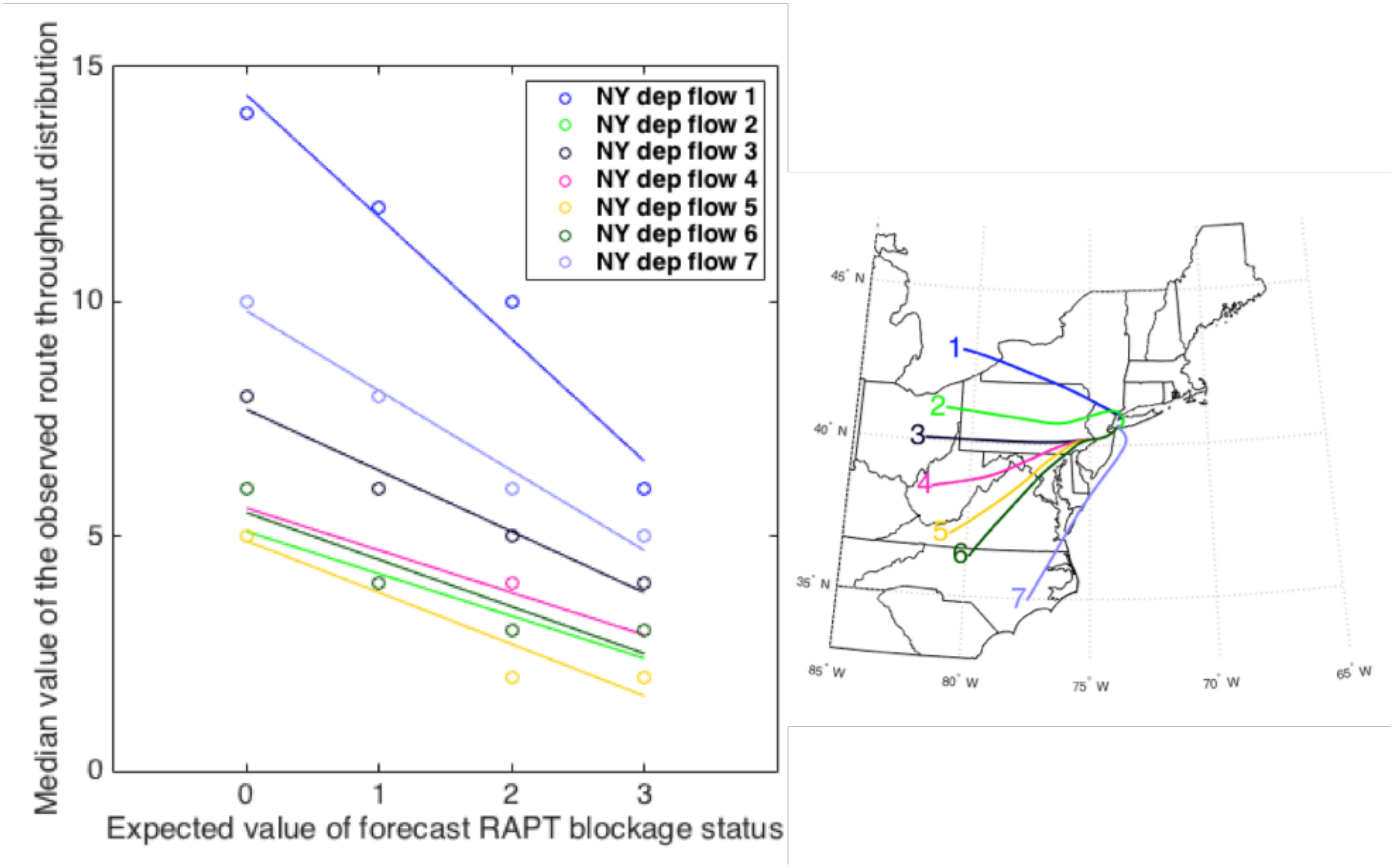
Resource Use Patterns

- Patterns are correlated with time of day and constraints in the system
  - Probability of RUP occurrence by levels of convective weather impact (RAPT) reveals interesting aggregate behaviors



Resource Use Patterns

- Detailed RUP characterization enables the quantification of throughput reductions associated with weather impacts at the route level and provides a foundation for the development of predictive airspace capacity models



Summary and Next Steps

- Analytics framework for characterizing air traffic flows based on historical radar tracks
- Application to the NY Metro
  - Identification of major trajectory patterns (under nominal and off-nominal conditions)
  - Assessment of trajectory conformance and identification of days with significant irregularity in operations
  - Identification and characterization of resource use patterns
  - Preliminary insights about how constraints imposed by convective weather impact system throughput in aggregate and individual route perspectives
- Development of predictive models of airspace capacity that capture the actual behavior of the system under different conditions can provide the basis for tactical traffic flow management mechanisms

Questions?

mayara@mit.edu